UC San Diego

UC San Diego Previously Published Works

Title

Identification of H3K4me1-associated proteins at mammalian enhancers.

Permalink

https://escholarship.org/uc/item/9s9286nj

Journal

Nature genetics, 50(1)

ISSN

1061-4036

Authors

Local, Andrea
Huang, Hui
Albuquerque, Claudio P
et al.

Publication Date

2018

DOI

10.1038/s41588-017-0015-6

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2018 June 19.

Published in final edited form as:

Nat Genet. 2018 January ; 50(1): 73–82. doi:10.1038/s41588-017-0015-6.

Identification of H3K4me1-Associated Proteins at Mammalian Enhancers

Andrea Local[1,#,*], **Hui Huang**[1,2,*], **Claudio P. Albuquerque**[1], **Namit Singh**[1], **Ah Young Lee**[1], **Wei Wang**[3], **Chaochen Wang**[3], **Judy E. Hsia**[1], **Andrew K. Shiau**[1], **Kai Ge**[4], **Kevin Corbett**[1,5], **Dong Wang**[3], **Huilin Zhou**[1,4,5], and **Bing Ren**[1,5,6]

[1]Ludwig Institute for Cancer Research, La Jolla, California [2]University of California San Diego, Biomedical Sciences Graduate Program, La Jolla, California [3]Skaggs School of Pharmacy & Pharmaceutical Sciences, University of California San Diego, La Jolla, California [4]National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland [5]University of California San Diego School of Medicine, Department of Cellular and Molecular Medicine, La Jolla, California [6]Center for Epigenomics, Institute of Genomic Medicine, La Jolla, California

Abstract

Enhancers act to regulate cell type specific gene expression by facilitating the transcription of target genes. In mammalian cells active or primed enhancers are commonly marked by monomethylation of Histone H3 at lysine 4 (H3K4me1) in a cell-type specific manner. Whether and how this histone modification regulates enhancer-dependent transcription programs in mammals is unclear. In this study, we conducted SILAC Mass-spec experiments with mono-nucleosomes and identified multiple H3K4me1 associated proteins, including many involved in chromatin remodeling. We demonstrate that H3K4me1 augments the association of the chromatin remodeling complex BAF to enhancers *in vivo* and that *in vitro*, H3K4me1 nucleosomes are more efficiently remodeled by the BAF complex. Crystal structures of BAF component BAF45c reveal that monomethylation, but not trimethylation, is accommodated by BAF45c's H3K4 binding site. Our results suggest that H3K4me1 plays an active role at enhancers by facilitating the binding of the BAF complex and possibly other chromatin regulators.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence: Bing Ren (biren@ucsd.edu).
[#]Present address – Aptose Biosciences Inc. San Diego, California
[*]These authors contributed equally to this work

Author Contributions

A.L. and B.R. conceived the study and prepared the manuscript. A.L. designed and carried out the SILAC experiments, nucleosome pull-down experiments, and ChIP-seq experiments, and prepared the manuscript. H.H. performed H3K4me2 ChIP-seq analysis and all experiments with the dCD cell lines. A.Y.L. prepared sequencing libraries. C.A. ran the mass spec samples in the laboratory of H.Z. and provided expertise in mass spec analysis. H.H. performed ChIP-seq data analysis. C.W. and K.G. provided KMT2C/D DKO mESCs and shared expertise and data. W.W. and D.W. designed and executed the remodeling assays. A.K.S. designed/supplied the H3 tail peptides and, along with J.E.H., provided advice on their use in biochemical studies. N.S. purified Baf45c, performed H3 tail peptide binding measurements, and determined crystal structures under the direction of K.D.C..

Competing financial Interests

The authors declare no competing financial interests.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In cells, cis-regulatory elements such as enhancers and promoters can be defined not only by DNA sequence motifs but also by common and predictive patterns of epigenetic modifications[1]. Active promoters are enriched for H3K4me3, H3/H4 acetylation along with binding of multiple chromatin regulatory complexes[2]. Primed enhancers are marked by H3K4me1 (coupled with a depletion of H3K4me3) whereas active enhancers are enriched for H3K4me1, H3K27ac and sometimes H4K16ac and H3K122ac[2–8]. Such epigenetic signatures are commonly used to predict de novo regulatory elements in novel cell types. Numerous studies have demonstrated that H3K4me1 is highly dynamic and correlates well with cell-type specific gene expression profiles, whereas promoter-associated H3K4me3 is more invariant across cell types[9].

It has been postulated that specific histone modifications function as binding elements for effector proteins that serve to regulate transcription through manipulation of the chromatin environment or assembly of transcription machinery[10–13]. For example, promoter-associated H3K4me3 can lead to recruitment of TFIID (through direct interaction with TAF3) to positively regulate transcription[14]. On the other hand, the function of H3K4me1 at enhancers has not been well understood. In Drosophila, knockout of the Trithorax-related (Trr) histone methyltransferase results in a global loss of H3K4me1[15] and a concomitant loss of enhancer function[15,16]. Similarly, loss of KMT2C/D, the human homologs of Trr, abolishes H3K4me1 and reduces H3K27ac levels as well as binding of Mediator and RNA polymerase II at enhancers[16,17]. KMT2C/D knockout cells exhibited defects in enhancer activation, cell type specific gene expression and differentiation capacities[15,17]. These studies, while supporting a role for H3K4me1 in enhancer function, did not reveal the mechanism of action by this histone mark. It is very likely that H3K4me1 may act by recruiting specific effector proteins.

A recent study of the H3K4 demethylase KDM5C revealed that while H3K4me3 positively regulates transcription at promoters, increased H3K4me3 serves to decrease enhancer function[18]. The correct balance of H3K4me1 and me3 at promoters is equally important for transcriptional regulation. At promoters, a decrease of H3K4me3 and repression of transcription is coupled with an increase of H3K4me1 in many cell types[19]. Additionally, H3K4me1 is known to block binding of H3K4me3-associated factors such as ING1. In fact, H3K4me1 also demarcates the boundaries of active promoters, thus limiting the recruitment of factors and specifying the promoter region[19]. Clearly these closely related modifications play very distinct roles in gene regulatory networks in cells, depending on localization and differential association with regulatory complexes. This fact underscores the need to identify factors that can specifically bind to H3K4me1, and perhaps distinguish between H3K4me1 and me3, in order to fully understand the role of this histone modification in gene regulation.

Peptide or nucleosome pulldown coupled with SILAC mass spec analysis has been utilized to identify factors associating specifically with histone tail modifications[14,20,21]. Such studies have successfully identified proteins associated with H3K9me, H3K4me3, and H3K27me3. However, in all previous studies, binding of complexes to methylated versus unmethylated histone states was compared. In the current study, we designed a screen to identify candidate H3K4me1 binders while simultaneously comparing association of factors with mononucleosomes bearing the H3K4me1 versus H3K4me3 modification. Our approach identified multiple components of the transcriptional regulatory machinery including the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

BAF complex as enriched for H3K4me1 association. ChIP-seq analysis confirmed that these factors' binding to putative enhancers correlates with H3K4me1 genome-wide in mESCs. Importantly, binding of these H3K4me1-associating proteins was drastically reduced upon depletion of KMT2C/D and loss of H3K4me1 at enhancers. In addition, loss of H3K4me1 in a mutant mouse ES cell line bearing catalytic site mutations in KMT2C and KMT2D correlates with reduced binding of BAF components SMARCA4 (BRG1) and DPF2 (BAF45d). We characterized the subunit in the BAF complex involved in preferential recognition of H3K4me1 over H3K4me3 by X-ray crystallographic analysis. We further demonstrated that *in vitro* BAF more efficiently remodels H3K4me1 nucleosomes. Taken together, our results provide mechanistic insights by which H3K4me1 acts to regulate the function of enhancers.

Results

Identification of potential H3K4me1 binding partners

We assembled nucleosomes with chemically modified histone H3 and naïve H4, H2A, and H2B (Fig. 1A)[22–25]. The H3K4me1 and H3K4me3 nucleosomes were used as baits in pulldowns from nuclear extract (NE) prepared from HeLa cells grown in media containing either light or heavy isotope-labeled amino acids as shown in Figure 1B[20]. Any factor specifically associating with H3K4me1 over H3K4me3 in the forward reaction would be detected by mass spec as enriched in light Lys labeled peptides, and in heavy Lys labeled peptides in the reverse reaction (Fig. 1B). Multiple replicates were performed with similar results. For final analysis, 2 replicates were combined and ratios of light peptides to heavy peptides were averaged across replicates (Fig. 1C and Supplemental Table S1). As we are only assessing H3K4me1 vs me3 affinities we cannot rule out the possibility that factors identified as H3K4me1 binders may also associate with H3K4me2 or me0. Nevertheless, our approach yielded a plethora of putative H3K4me1 associated proteins including many known chromatin regulators and chromatin associated factors (Supplemental Table S2). Multiple subunits of the BAF (SWI/SNF) complex, such as SMARCA4 (BRG1) and SMARCC1/2 (BAF155/170), were isolated in the precipitates. Also identified were components of other chromatin remodeling complexes such as BAZ1B from WINAC and WICH, and BAZ1A from ACF. Many factors isolated contain histone-binding domains (Supplementary Table S2) and, in addition, several of these factors have been found associated with H3K4me1 regions of the genome in cells by ChIP mass spectrometry[26]. Interestingly, two Cohesin subunits were found to be associated with H3K4me1-nucleosomes. Cohesin is known to associate with enhancers and facilitate enhancer-promoter looping[27]. The results implicate H3K4me1 in many facets of enhancer function from chromatin remodeling to looping of enhancers and promoters. In addition to the H3K4me1 associated factors we identified several novel H3K4me3 associated proteins such as the FACT components SSRP1 and SUPT16H.

Our mono-nucleosome pulldowns differed from previous experiments that largely employed methylated histone tail peptides as bait. For the purpose of comparison, the assay was repeated comparing H3K4me1 and H3K4me3 peptides instead of mono-nucleosomes and in this case we observed enrichment of TAF and ING family proteins as observed by other

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

labs[14]. Notably, there was less enrichment of factors for H3K4me1 in the peptide pulldowns, compared to the use of mono-nucleosome templates. This difference could be due to histone tails adopting a distinct conformation, necessary for substrates to bind, only in the presence of intact nucleosomes[28]. Alternatively, it could be due to additional interactions that exist only in intact nucleosome substrates.

To validate association and identity of a subset of the chromatin regulators (CRs) identified in our screen we incubated methylated nucleosomes with HeLa NE and performed western blotting to identify associated factors (Fig. 1D). Target validation was limited by availability of specific antibodies so unfortunately we were unable to conduct further analysis on several interesting candidates. However, we confirmed preferential binding of H3K4me1 over H3K4me3 by a number of known enhancer-associated factors. It should also be noted that some proteins bind to multiple methylation states, such as Sap18 to H3K4me1/me2 and SMARCC2 to H3K4me0/1 (Fig. 1D). While some factors have domains known to bind methylated Lysine residues, such as PHD domains found in PHRF1, and BAF components, other factors identified in the screen do not have any know histone binding domains. It is clear that complex binding patterns of multiple protein complexes is involved.

CRs are localized to H3K4me1 rich regions of the genome

Next we performed ChIP-seq for 16 CRs and 4 histone modification marks in mouse embryonic stem cells (mESCs) to determine the localization of the candidate H3K4me1-binding chromatin regulators (CRs). Clustering analysis of the ChIP-seq profiles of these factors along with three histone H3 lysine 4 methylation states (me1, me2 and me3) showed that nearly all of the CRs tested cluster together with H3K4me1 in a branch separate from H3K4me2 and H3K4me3 (Fig. 2A). We further assayed the binding of the CRs to a subset of previously validated enhancers[29] and negative control regions by ChIP-qPCR, and found CRs to be enriched at all enhancers tested (Fig. 2B–C and Fig. S1A–D). Enrichment of H3K4me1-associated CRs was observed at a previously validated Sox2 enhancer[30], and several factors are also enriched at the Sox2 promoter overlapping with the promoter-flanking H3K4me1 domains. Interestingly we observed consistently higher CR enrichment at regions with both H3K4me1 and H3K27ac (Fig. S2C and S2D). Next we investigated CR association with poised (n=28,008) and active (n=13,811) enhancer regions, defined as H3K4me1-positive regions with or without concomitant H3K27ac signals. For this specific analysis “active” enhancers were defined based on H3K27ac signals and not H3K16ac or H3K122ac. We discovered that active enhancer regions tend to be occupied by multiple CRs while poised enhancer regions show individual CR binding patterns (Fig. 2D and Fig. S2D left vs right panel). The majority of CRs tested bound a high fraction of H3K27ac containing enhancer regions (Fig. S1E). That acetylation of H3K27 at enhancers coincides with binding by multiple co-activators implies that binding of multiple CRs might be necessary for full activation of the enhancers.

H3K4me1 dependent association of CRs with enhancers

The above results confirmed the association of CR complexes to H3K4me1 *in vitro* and *in vivo*. To determine if chromatin association of CRs is dependent upon H3K4me1, we carried out ChIP-seq analyses of these protein complexes in mouse ESC deleted of KMT2C/D[31].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Previous studies have demonstrated that KMT2C/D are responsible for H3K4me1 deposition at enhancers in multiple species[15–17]. Consistent with previous data from mouse pre-adipocytes and human colon cancer cells, knockout of both of these enzymes in mouse ESCs results in a general decrease in H3K4me1 but has little effect on the global level of H3K4me3[31]. We performed H3K4me1, H3K4me2, and H3K4me3 ChIP-seq in mESCs deleted of both KMT2C and KMT2D genes (DKO) and compared the results with the data from WT mESCs. We observed that the majority of H3K4me3 distribution remains unaltered between WT and DKO (Fig. 3C) whereas H3K4me2 levels are mildly effected (Fig. 3B and Fig. S2A–B). Consistent with the previous studies we observed a dramatic reduction in H3K4me1 signal throughout the genome (Fig. 3A and Fig. S2A–B): 47% of H3K4me1 peaks detected in WT mESCs were lost in DKO mESC (Fig. 3D, Fig. S2C). The KMT2C/D-dependent H3K4me1 peaks are enriched at enhancers (Fig. 3D), consistent with previously suggested function of KMT2C/D at these sites[15–17]. KMT2C/D-independent H3K4me1 peaks, on the other hand, overlap not only with enhancers and but also promoters (Fig. 3D). We also detected both KMT2C/D dependent and independent H3K4me2 peaks (Fig. S2D). However, in contrast to H3K4me1 peaks, the KMT2C/D-dependent-H3K4me2 is found at both enhancers and promoters at equal proportions. Additionally, as seen in pre-adipocytes, KMT2C/D dependent loss of H3K4me1 also coincides with a moderate decrease in H3K27ac at the same regions (Fig. 3E and Fig. S2A and S2B).

Both KMT2C/D dependent and independent peaks are bound by CRs but the fraction of associated peaks is highly variable (Fig. 3F). CRs should be reduced at KMT2C/D dependent sites in DKO cells if H3K4me1 acts to facilitate or stabilize their binding. To test this hypothesis, we performed ChIP-seq for a subset of the H3K4me1-associated CRs and demonstrate and overlap with H3K4me1 occupancy in the wild-type cells. All CRs tested were reduced at KMT2C/D dependent H3K4me1 sites compared to KMT2C/D independent sites in the DKO mESCs (Fig. 3F and S2). We obtained similar results assessing CR association with known mESC enhancers using ChIP-qPCR (Fig. S2E).

A recent study by Dorighi and colleagues highlights a role for KMT2C/D in transcription regulation independent of H3K4me1 deposition[32]. Our data suggests that H3K4me1 is important for CR binding, however this new study raised the possibility that loss of KMT2C/D could directly affect binding of CRs independently of H3K4me1 loss. We therefore utilized the KMT2C/D catalytically inactive cell line (dCD) to distinguish between the role of H3K4me1 and KMT2C/D in binding of CRs. We performed ChIP-seq for H3K4me marks, H3K27ac, and BAF complex components SMARCA4 (BRG1) and DPF2 (BAF45d) (Fig. 4A). In the dCD cells 38% of the distal H3K4me1 sites had reduced levels of H3K4me1. Interestingly, a small fraction of H3K4me1 sites also gained H3K4me1 signal, which is consistent with the previous data[32], and these sites are located closer to promoters than the H3K4me1 depleted regions. As in DKO cells, H3K4me2 and me3 levels were less affected than H3K4me1 (Fig 4B–C, S3B). At regions where we observed specific loss of H3K4me1 signal we likewise observed a decrease in binding of both SMARCA4 and DPF2 (Fig 4E–F, S3C–D). Reduced BAF complex binding is specific for sites where H3K4me1 is depleted (Fig. 4F) and was not seen at sites where H3K4me1 is unchanged, confirming the role of H3K4me1 in facilitating BAF binding to these regions. Taken together, our data from

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

KMT2C/D KO and catalytically inactive cells supports the hypothesis that H3K4me1 plays an important role in binding of multiple CR complexes to enhancers.

BAF complex preferentially binds to and remodels H3K4me1 nucleosomes

The BAF complex is known to co-localize with H3K4me1 in the genome[6]. Our data suggests that H3K4me1 may play a direct role in stabilizing BAF complex binding to chromatin. To confirm that H3K4me1 can indeed serve to facilitate binding of BAF complexes in the absence of other co-factors or transcription factors, we repeated the mono-nucleosome pulldown assays with BAF complex purified from HeLa cells (Fig. S4A). We demonstrate that purified BAF complex binds to H3K4me1 with higher affinity than H3K4me3 on mono-nucleosomes (Fig. 5A) and, to a lesser extent, H3 tail peptides (Fig. S4B). These data demonstrate that protein complexes can recognize and distinguish between closely related H3K4 methylation states, and this could be important for their recruitment to enhancers. The BAF complex regulates transcription by remodeling nucleosomes at sites of H3K4me1, suggesting a link between histone methylation and BAF activity. Utilizing *in vitro* nucleosome remodeling assays[33] we find that the BAF complex more efficiently remodels H3K4me1 mono-nucleosomes, than H3K4me0, H3K4me2, and H3K4me3 mono-nucleosomes (Fig. 5B–C, S4C). This data suggests a functional link between enhancer-specific histone modifications and the activity of recruited chromatin regulatory complexes.

Crystal structure of DPF3 binding preferentially H3K4me1

Based on peptide binding and NMR/X-ray structures, the PHD1 domain of BAF component DPF3 (BAF45c) recognizes H3K14ac, while the PHD2 domain in these proteins binds to H3K4me0[34]. BAF subunits DPF1, DPF2, DPF3, and PHF10 (BAF45B, C, D, and A isoforms respectively) have cell type specific expression patterns[35]. Our data demonstrates that mESC specific DPF2 associates with H3K4me1. To determine if the DPF3 (BAF45c) PHD2 domain could contribute to H3K4me1 recognition as well, we purified the PHD1/2 region of DPF3 of this family of proteins, and used isothermal titration calorimetry to measure its affinity for H3 tail peptides containing H3K14ac plus H3K4me0, H3K4me1, or H3K4me3. Consistent with our biochemical studies, we found that the isolated BAF45c PHD1/2 region strongly preferred H3K4me1 (K_d of 20 μM for H3K4me1/K14ac) over H3K4me3 binding (K_d of 115 μM for H3K4me3/K14ac) (Fig. S5A–C). However, in contrast to our findings with the intact BAF complex and mono-nucleosomes the DPF3 PHD1/2 region bound to the H3K4me0 peptide with slightly higher affinity (K_d of 7.8 μM for H3K4me0/K14ac) than the H3K4me1 peptide. These data suggest that additional factors in the BAF complex and/or nucleosomes may influence H3K4me1 specificity.

To reveal the atomic basis of the preferential recognition of DPF3 PHD1/2 for H3K4me1 over H3K4me3, we next determined two high-resolution (1.2 Å) crystal structures of the DPF3 PHD1–2 region bound to H3 tail peptides (residues 1–18) containing H3K14ac and either H3K4me0 or H3K4me1 (Supplemental Table 3). The two structures show a nearly identical overall structure of DPF3 (<0.04 Å overall Cα r.m.s.d.), and largely agree with prior structures of this protein, with a 1.5 Å overall Cα r.m.s.d. to a prior NMR structure (PDB ID 2KWJ) and 0.8 Å overall Cα r.m.s.d. to a prior X-ray crystal structure (PDB ID 5I3L)[31,34]. In our two structures, the two PHD domains are intimately associated with one

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

another, with a binding pocket in PHD1 that recognizes H3K14ac, and a pocket in PHD2 that recognizes H3K4 (Figure 6A–C) leading to virtually identical bound conformations of the H3K4me0 and H3K4me1 peptides. In both complexes, H3K4 is nestled tightly in a surface cavity made up of the hydrophobic side chains of I314, L331, and F333. In addition, the main chain carbonyl groups of residues 314, 315, and 317 are all close enough to the H3K4 amino group to form hydrogen-bonding interactions. These interactions likely contribute to the preferential binding of unmethylated or monomethylated H3K4, the amino groups of which can form two (K4me1) or three (K4me0) hydrogen bonds, over di-or trimethylated H3K4. In addition, the H3K4 mono-methyl group packs in a preformed cavity that is just large enough for a single methyl group. Hence, these carbonyls may sterically disfavor di- or trimethylated H3K4 binding.

In contrast to earlier NMR structures of the DPF3-H3 tail complex[34], but in agreement with a recent crystal structure[31], our structures show that H3 residues 4–10 adopt an α-helical conformation. Additionally, we find that H3R8 forms a “lid” over the binding site, extending directly over H3K4 and forming a hydrogen-bonding network with DPF3 residues E315 and D328 on opposite sides of the H3K4 binding pocket (Fig. 6B–C and S5D–E); this residue’s position was not well-resolved in the previous crystal structure[31]. Both the α-helical conformation of the H3 tail and the H3R8 “lid” most closely mirror earlier observations in crystal structures of the MYST family acetyltransferase KAT6A (MOZ), which possesses a double-PHD finger domain at its N-terminus that recognizes unmodified H3K4 and acetylated H3K14[28] or propionylated/butyrylated/crotonylated H3K14[36]. This H3 tail-binding mode may also be shared in other double-PHD finger protein families; for instance, an unpublished NMR structure of KMT2C (PDB code 2YSM) shows that this protein possesses a pair of acidic residues bracketing the H3K4 binding site that could participate in H3R8 binding. This mode of H3K4 recognition may also have functional relevance as it leaves the H3K4me1 group solvent exposed in the complex, creating the possibility that additional factors in BAF or in the nucleosome itself could associate with the composite DPF3-H3K4me1 surface and provide additional specificity for H3K4me1 over H3K4me0.

Discussion

In summary, we carried out SILAC mass spectrometry analysis to systematically identify nuclear proteins that bind H3K4me1. Our experiments uncovered components of multiple chromatin regulatory complexes, including the BAF chromatin remodeling complex, as H3K4me1-associating proteins. We further validated the binding of a subset of these complexes to H3K4me1 mononucleosomes *in vitro* and to genomic regions bearing the histone mark in embryonic stem cells. We showed that deletion of H3K4 methyltransferaseses KMT2C/D leads to a loss of occupancy by these complexes at KMT2C/D-dependent H3K4me1 regions. Importantly, we confirmed that loss of H3K4me1 in both KMT2C/D knock out and catalytically null mutant cells correlated with a decrease in binding of CRs to enhancers, supporting our hypothesis that H3K4me1 plays an important role in binding of key chromatin regulatory factors. We chose to focus on the BAF complex, and obtained strong evidence suggesting that H3K4me1 is directly involved in the association of this complex to chromatin. The BAF complex belongs to the SWI/SNF family of ATP-dependent chromatin remodeling complexes[35]. Containing between 10 to 12

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

components, BAF complexes are necessary for early embryogenesis, activation of lineage specific genes during cellular differentiation, and maintenance of pluripotency of embryonic stem cells. Genome-wide profiling studies have shown that BAF complexes generally localize to distal enhancers where they are required for histone acetylation during differentiation of ES cells. A recent study involving *in situ* capture of specific genomic regions also identified BAF as an enhancer bound complex[37]. However, exactly how BAF complex is recruited to the enhancers has not been fully understood[35]. Here, we provided multiple lines of evidence that H3K4me1 may play a role in the recruitment of BAF complex to enhancers. BAF complexes fail to localize to promoter-distal enhancers in KMT2C/D double-KO, and in KMT2C/D catalytically inactive mutant cells. Using protein-pull down assays, we showed that the BAF complex interacts directly with H3K4me1 mononucleosome *in vitro* via the PHD2 domain in DPF3 (BAF45c). X-ray crystallography experiments further revealed a surface cavity in the PHD2 domain of DPF3 that readily accommodates monomethylated lysine 4 of histone H3, but not tri-methylation. Finally, nucleosome remodeling assays demonstrated that H3K4me1 facilitates the BAF complex' nucleosome remodeling activity above all other H3K4me states. These results, taken together, support a model in which the histone modification H3K4me1 directly helps to recruit BAF complex to enhancers, and therefore plays an active role in enhancer function.

While this work was under revision, Dorighi and colleagues[32] reported that KMT2C/D promotes RNA synthesis at enhancers and nearby promoters independently of the H3K4 monomethylation activities. While this observation suggests that H3K4me1 may not be necessary for loading of RNA polymerase II at enhancers and subsequent activation of target promoters, it does not rule out other functions of H3K4me1 at enhancers. Another recent study demonstrated that Drosophila bearing catalytically inactive Trr (H3K4me1 histone methyltransferase) survive to adulthood with only subtle gene expression changes. However, if subjected to temperature stress conditions developmental abnormalities were observed[38]. In addition, this and other studies have found that loss of KMT2C/D in mESCs does not affect self-renewal[17,38]. This can be partially explained by the fact that at poised enhancers in mESCs H3K4me1 is KMT2C/D independent[32], suggesting a role for other methyltransferases in H3K4me1 deposition and enhancer function in higher organisms. This is in agreement with our current study demonstrating that ~50% of H3K4me1 peaks in mESCs are KMT2C/D independent. Therefore, additional experiments are needed to better define the role of H3K4me1 in enhancer function during cellular differentiation and animal development.

Online Methods

Protein Purification

Individual histones were expressed in *E.coli* as described by K. Luger[1,2]. We utilized wild type H2B, H2A, H3, H4 and a histone 3 C110A K4C mutant construct (provided by Dr. M. Carey, UCLA). Methyl-lysine analogs were made as previously described[3]. Briefly, 5 mg of lyophilized H3 was mixed with (2-hal-oethyl) amines under reducing conditions. After quenching with β-mercaptoethanol the resultant methylated histone was dialyzed against water overnight, spun to remove precipitant, aliquoted and lyophilized. Histones were stored

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

lyophilized at −80°C until used. Equimolar amounts of histones were combined under denaturing conditions and then dialyzed overnight to assemble octamers that were then purified via size selection[2]. Octamers were stored at 4°C or aliquoted at −80°C. Biotin tagged 601λ positioning sequence was prepared as described[4], mono-nucleosomes were produced via salt dialysis[5]. We tested lysine methylation by western blotting with antibodies validated to recognize specific H3K4me states. The resulting mono-nucleosomes were immobilized to streptavidin coated beads (Invitrogen MyOneT1 cat# 65602) as per manufacturers binding instructions and employed as a bait in a variety of binding studies.

BAF complex was purified from HeLa cells stably expressing F-Ini1 (BAF47). HeLa NE was prepared using standard conditions and buffers[6,7]. Magnetic Flag M2 beads (Invitrogen cat#M8823) were washed 3 times with 140mM NaCl, 25mM Tris pH 8.0, 1mM EDTA then incubated for 2 hours at 4c with Flag-Ini1 NE. Beads were washed 3 times with 250mM NaCl buffer and Flag-BAF complex eluted in above buffer plus 1mM DTT and 20 fold excess of Flag peptide (Sigma Cat#F3290). The BAF complex was validated by silver stain, and western blotting with Flag M2 antibody.

We cloned the PHD1-2 region of human BAF45c (residues 254–368) into a pET3a-derived vector containing an N-terminal His_6-SUMO tag, and expressed in *E. coli* Rosetta DE3 pLysS (EMD Millipore) in 2XYT media supplemented with 0.1 mM $ZnCl_2$. We grew cells to an OD_{600} of 0.6 at 37°C, cooled to 18°C and induced with 0.3 mM IPTG for 16 hours. We purified the protein using Ni^{2+} affinity purification (Qiagen Ni-NTA Superflow), followed by cleavage of the tag with TEV protease[8] and size-exclusion chromatography (Superdex 200, GE Life Sciences). We concentrated the purified protein to 20 mg/mL in a buffer containing 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2mM DTT, and generated peptide complexes by mixing peptides – H3K14Ac (ARTKQTRARKSTGGK(Ac)APRK) or H3K4me1K14Ac (ARTK(me)QTRARKSTGGK(Ac)APRK) – with BAF45c at a 1.2:1 molar ratio, and incubating at room temperature for 30 minutes.

SILAC Mass Spec

HeLa cells (ATCC CCL-2) were grown in normal DMEM culture media, with 10% FBS and 1% Pen/Strep (termed light from here on) or in culture media containing N15 and C13 labeled Arginine and Lysine (termed heavy NE). NE was prepared using a small scale Digman and Roeder protocol[7,9]. Concentration of NE batches was determined by BCA assay as per manufacturer's instructions (Pierce cat# 23228). For the forward reaction 25 μg of immobilized H3K4me1 mononucleosomes was incubated with 1mg of "light" NE and in a separate tube 25 μg of H3K4me3 nucleosomes were incubated with an equal amount of "heavy" NE ($^{13}C_6$,$^{15}N_4$-L-arginine and $^{13}C_6$,$^{15}N_2$-L-lysine) at 4°C for 4 hours. Beads were washed 3 times in 250mM NaCl, 25mM Tris pH 8.0, 1mM EDTA, 1mM DTT, and 0.2% NP40. Bound factors were eluted in 400 μl of 4M Urea, 500mM NaCl, 20mM HEPES pH 8.0, and 0.2% NP40. Eluates were combined, alkylated, and digested by trypsin. Digested peptides were desalted by C18, fractionated by HILIC column (Tosoh Bioscience Cat # 21486), and analyzed by LC-MS/MS using an Orbitrap-LTQ mass spectrometer (MS)[10]. MS data was then searched on Sorcerer-SEQUEST using a non-redundant human database downloaded from Uniport with 50ppm parental mass tolerance and a 0.8 cutoff for the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

peptide prophet probability used to filter the dataset similar to that previously described[11]. The identified peptides were quantified using XPRESS with a 50ppm mass tolerance and ± 25 scans from the apex. The median of the peptide ratio was then calculated. The reverse reaction was set up with the same concentrations but with H3K4me1 with heavy extract and H3K4me3 with light extract. Replicate reactions were performed with different batches of HeLa NE and chemically modified histones. For the forward reaction we calculated the ratio of light to heavy peptides and for the reverse reaction the ratio of heavy to light peptides and logarithm-transformed ratios were plotted. The top right quadrant includes any proteins enriched for H3K4me1 and the bottom left quadrant includes those enriched for H3K4me3 binding. Replicate pulldowns performed with NE and modified nucleosomes prepared as independent biological replicates. As expected, common processing contaminants (such as actin, ribosomal protein and keratin), along with histones that were part of the nucleosomes used in the purification were removed from the final data sets listed. Proteins in the final dataset were required to be found in both the forward and reverse reactions and their enrichment for both mono- and tri-methylation was required to be consistently positive or negative. Proteins that did not meet this criteria were removed from the dataset. SILAC experiments were performed twice with two separate chemically modified nucleosome sets and biological replicate nuclear extracts.

Methyl-nucleosome Pulldowns

3 μg of mono-nucleosomes were pre-bound to MyOneT1 beads and binding was verified by Western blotting with H2B antibody. Immobilized nucleosomes were incubated with HeLa NE (200 μl of ~5 mg/ml), GST-PHD domain (2 μg) or Flag-tagged BAF complex for 1 hour at room temperature. Purified domains or complexes were bound in 140mM NaCl, 25mM Tris pH 8.0, 1mM EDTA, and 100mM ZnCl2. Beads were washed 3 times with 250mM NaCl, 25mM Tris pH 8.0, 1mM EDTA, 0.2% NP40, and 1mM DTT and resuspended in equal volume of 2X Laemmli Sample Buffer (Biorad cat# 161-0737). Binding was assayed via western blotting with antibodies listed in Supplemental table 4. At least 2 biological replicates were performed for each pulldown. Multiple factors were probed on each membrane, uncropped original images are shown in Supplementary figure 6.

Nucleosome Remodeling Assay

The 216-bp DNA fragment containing 146-bp Widom601 sequence in the middle was P^{32} labeled by PCR amplification from the pGEM-3Z-601 vector[12]. Four types of human histone octamer (wt, H3K4me1, H3K4me2, and H3K4me3) were ordered from EpiCypher, *Inc.*[13]. Mono-nucleosomes were assembled onto the labeled template by salt serial dilution method as previously described[14]. Cold mono-nucleosome (2 μM) was assembled using *Xenopus laevis* histones with unlabeled DNA fragment by the same method, which is then mixed with sonicated calf thymus DNA (1mg/ml, Sigma). 1 μl purified BAF complex (0.15 mg/ml) was incubated with 2 μl P^{32} labeled mono-nucleosomes (25 nM) at room temperature in final 10 μl remodeling buffer (20 mM Tris-HCl pH7.5, 50 mM NaCl, 2.5 mM $MgCl_2$, 2 mM DTT, 100 μg/ml BSA, 5% Glycerol, 0.01% NP-40, 0.01% Triton X-100, and 1 mM ATP). After 30min, 1.5 μl cold nucleosome was added to quench the reaction for 30 minutes at room temperature. The reaction products were loaded onto 5% native polyacrylamide gels and resolved by electrophoresis at 4°C for 2.5 hours at 200 volts. Gels

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

were exposed to storage phosphor screens and scanned with a Typhoon imaging scanner. Quantified calculations were done using the software of Image Lab 4.1 (Bio-rad). Error bars, mean ±SD n=4 replicates.

Chromatin Immunoprecipitation

KMT2C/D DKO cells and control WT mESC lines were previously described[15]. Cells were cultured on MEF feeder cells in serum free media (DMEM 10013-cv supplemented with 15% KSR, 2 mM L-glutamine, 100 μM nonessential amino acids, 50 μM beta-mercaptoethanol, 1000 units ml^{-1}; leukemia inhibitory factor (LIF), 100 units ml^{-1} penicillin, and 100 μg ml^{-1} streptomycin.) at 37°C with 5%CO2. ChIP was performed as previously described[16,17] with antibodies listed in Supplemental Table 5. Briefly, 2 μg of specific antibodies were bound to Dynal magnetic secondary beads for 6 hours. After washing, 20–50 μg of sonicated chromatin from WT or KMT2C/D DKO mESCs was incubated with the bead bound antibody overnight. mESCs previously validated in Mouse model Encode studies. Beads were then washed and enriched DNA purified. The ChIP DNA was then sequenced as previously described or analyzed by qPCR. For qPCR, we utilized Sybr green 1 master mix (Roche cat# 04707516001) and mESC specific enhancer primers (Supplemental Table 5). Reactions were performed in triplicate and data shown is average of at least 3 biological replicates, error bars represent one standard deviation from mean.

Sequencing Read Alignment

ChIP-seq libraries were sequenced on the Hi-Seq 2500 platforms. 36bp single end sequencing data was aligned to a reference mouse genome (mm9) downloaded from UCSC genome browser by using Bowtie[18]. Unmapped and non-uniquely mapped reads were removed, and PCR duplicate reads were removed with Picard. ChIP-seq data sets are available on GEO accession number GSE80049.

Peak Identification

H3K4me1 and H3K4me3 peaks were defined by using MACS2 with default parameters except following options “–m 5 50 –p 1e-5” for each biological replicate. After that we selected the best quality replicate for each modification based on number of identified peaks. From the best quality replicate, we only selected reproducible peaks at least two biological replicates. H3K4me1 and H3K4me3 peaks were merged together and classified into H3K4me1 peaks, H3K4me3 peaks, and common H3K4me1/me3 peaks.

Chromatin regulator (CR) peaked regions were defined by comparing predefined H3K4me1/me3 peaked regions. Due to the limited quality of CR ChIP-seq results we first calculated input normalized CR ChIP-seq RPKMs on H3K4me1/me3 peaked regions for each biological replicate. If CR ChIP-seq RPKMs is enriched more than 1.5 fold compared to input RPKMs at least two biological replicates the h3k4me1/me3 peaked region was defined as occupied by the corresponding CR.

H3K27ac ChIP-seq data was processed similarly as described in CR ChIP-seq data analysis except using 2-fold enrichment threshold when we define H3K27ac peaked regions. The

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

density of histone modifications and CR binding was defined by taking average input normalized RPKMs of reproducible peaks across multiple biological replicates.

Clustering of CR-binding patterns

K-means clustering was performed for CR binding patterns according to poised enhancer and active enhancer regions. The peaks occupied by H3K4me1 but depleted by H3K4me3 were considered as enhancer regions after excluding upstream and downstream 2.5kb from RefSeq transcription start sites. Active enhancers were defined if the enhancer regions were overlapped with H3K27ac peaked regions (n=13,811). Otherwise the region was defined as poised enhancer regions (n=28,008). We performed K-mean clustering based on 13 CR binding patterns at active and poised enhancer regions separately, and generated 15 different clusters. Based on the enrichment of CR binding patterns in each cluster we manually assigned each cluster as ‘No CR bind’, ‘CR-specific bind’, and ‘Multiple CR bind’.

Identification of KMT2C/D dependent and independent H3K4me1 and H3K4me3 peaks

We defined KMT2C/D dependent and independent sites based on H3K4me1 levels between WT and KMT2C/D DKO cell lines. If the region is occupied by H3K4me1 more than 2-fold compared to input data and depleted more than 2-fold in KMT2C/D DKO cell lines the region was defined as a KMT2C/D dependent site, otherwise defined as a KMT2C/D independent site.

Analysis of CR-binding patterns in KMT2C/D DKO cell lines

To determine whether CR binding patterns are dependent on H3K4me1, we performed additional ChIP-seq for CRs in with KMT2C/D DKO cell lines. Sequence read alignment was performed as described in WT CR ChIP-seq data. During downstream analysis of DKO cell lines we only considered CR ChIP-seq peaked regions defined in WT.

Analysis of BAF complex association in KMT2C/D catalytically null mutant mES cells (dCD)

The KMT2C/D catalytically null mutant cells (dCD) and control WT mESC cell lines were a gift from Wysocka lab and described before[19]. Antibodies used were: H3K4me1 (Abcam, ab8895), H3K4me2 (Active Motif, 36979, clone 0303), H3K4me3 (Millipore, MC315,04-745), H3K27ac(Active Motif, 39685), BRG1 (Abcam, ab110641), DPF2 (Abcam, ab134942). The sequencing reads were aligned to mm9 (July 2007, NCBI37) by bwa (version 0.7.13-r1126). Unmapped, multimapped and PCR duplicates were removed by Picard (2.11.0). ChIP seq peaks were called for each replicate and pooled data using MACS2 (2.1.0.20150731) with default parameters. Pooled peaks found in both replicates were defined as reproducible peaks for each factor. ChIP Fold enrichment over input was calculated by *macs2 bdgcmp*. Normalized ChIP signal was generated by UCSC tool *bedGraphToBigWig*. Averaged ChIP-seq fold enrichment over input at each reproducible peak was calculated using *bigWigAvgOverBed.* Distal regions are peaks whose center are ± 2kb away from GENCODE transcript TSS. Using two fold difference cut off, distal regions were classified as “decreased”,”unchanged” and “increased” by comparing the fold enrichment over input in dCD to that in WT. If the fold enrichment of a region in dCD is less

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

than half of that in WT, the region was defined as “decreased” and *vice versa*. Heatmap and aggregate profile of input normalized ChIP-seq signal were generated by deepTools 2.5.1.

Isothermal Titration Calorimetry

Isothermal titration calorimetry was performed at the Sanford Burnham Prebys Medical Discovery Institute, Protein Analysis Core Facility. Assays were performed in buffer containing 20 mM Tris-HCl, pH 7.5, 300 mM NaCl and 1 mM DTT, using an ITC200 calorimeter from Microcal (Northampton, MA). Nineteen 2.0-μl aliquots of solution containing 1.0 mM peptides – H3K14Ac (ARTKQTRARKSTGGK(Ac)APRK), H3K4me1K14Ac (ARTK(me)QTRARKSTGGK(Ac)APRK), or H3K4me3K14Ac (ARTK(me3)QTRARKSTGGK(Ac)APRK) – were injected into the cell containing 200 μL of 80 μM BAF45c (residues 254–368) at 23°C. Data were analyzed using Origin software provided by Microcal. Peptides were synthesized by WuXi AppTec. Single replicate assay performed for each peptide.

Crystallography

We obtained BAF45c:peptide crystals by mixing 1:1.2 in a crystallization buffer containing 1.2 M tri-sodium citrate, 0.1 M HEPES pH 7.5 and 4% MPD. Despite growing crystals in similar conditions to those reported for an earlier structure of BAF45c bound to H3 tail peptide[15], we obtained crystals in a different space group ($P4_32_12$ versus $C222_1$). We flash-froze crystals without additional cryoprotection in liquid nitrogen, and collected diffraction data on beamline 9-2 at the Stanford Synchrotron Radiation Lightsource (support statement below). We processed all datasets with the SSRL autoxds script, which uses XDS[20] for data indexing and reduction), AIMLESS[21] for scaling and TRUNCATE[22] for conversion to structure factors. We determined the structure by SAD methods, with a 1.55 Å-resolution single-wavelength dataset collected at the Zn anomalous absorption peak. The Phenix Autosol pipeline[23] (HySS[24] for heavy-atom search, PHASER[25] for phasing, RESOLVE[26,27] for density modification and autobuilding) generated an initial model, which we manually rebuilt in COOT[28]. We refined this model against high-resolution (1.2 Å) datasets for each peptide (with a consistent free-R test set) in phenix.refine[29], using positional and individual anisotropic B-factor refinement for all non-hydrogen atoms, then iteratively rebuilt and refined each model. Both final models are of high quality as judged by refinement statistics and model geometry (Supplemental Table 3).

SSRL Support Statement—Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIGMS or NIH.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Data Availability and Accession Code Availability

The mass spectrometry proteomics data presented in figure 1 have been deposited to the ProteomeXchange Consortium (http://www.proteomexchange.org) with the dataset identifier PXD007942. ChIP-seq data sets used for analysis presented in figures 2, 3, 4, S1, S2, and S3 are available on GEO - accession number GSE80049. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=uhsroacglvcvvib&acc=GSE80049 Crystal structures presented in figure 6 and S5 are available through PDB https://www.ncbi.nlm.nih.gov/Structure/pdb and (ID: 5SZB and 5SZC) and the structural biology database (https://data.sbgrid.org/dataset/359/). Customized code for ChIP-seq analysis is available via GitHub. https://github.com/HuiHuang001/h3k4me1_associated_proteins_scripts

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Samantha Kuan and Bin Li for processing of ChIP-seq samples, Jason Liang and Gary Hon for help and advice in SILAC mass spec analysis, Drs. Joanna Wysocka and Kristel Dorighi for sharing the MLL3/4 dCD mESC line, and Inkyung Jung for advice on ChIP-seq data analysis. We also thank Timothy Gahman for arranging for peptide synthesis and Andrey Bobkov for assistance with isothermal titration calorimetry.

References

1. Hardison RC, Taylor J. Genomic approaches towards finding cis-regulatory modules in animals. Nat Rev Genet. 2012; 13:469–83. [PubMed: 22705667]
2. Heintzman ND, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007; 39:311–8. [PubMed: 17277777]
3. Creyghton MP, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010; 107:21931–6. [PubMed: 21106759]
4. Heintzman ND, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 2009; 459:108–12. [PubMed: 19295514]
5. Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol. 2015; 16:144–54. [PubMed: 25650801]
6. Rada-Iglesias A, et al. A unique chromatin signature uncovers early developmental enhancers in humans. Nature. 2011; 470:279–83. [PubMed: 21160473]
7. Taylor GC, Eskeland R, Hekimoglu-Balkan B, Pradeepa MM, Bickmore WA. H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. Genome Res. 2013; 23:2053–65. [PubMed: 23990607]
8. Pradeepa MM, et al. Histone H3 globular domain acetylation identifies a new class of enhancers. Nat Genet. 2016; 48:681–6. [PubMed: 27089178]
9. Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? Mol Cell. 2013; 49:825–37. [PubMed: 23473601]
10. Musselman CA, Lalonde ME, Cote J, Kutateladze TG. Perceiving the epigenetic landscape through histone readers. Nat Struct Mol Biol. 2012; 19:1218–27. [PubMed: 23211769]
11. Seet BT, Dikic I, Zhou MM, Pawson T. Reading protein modifications with interaction domains. Nat Rev Mol Cell Biol. 2006; 7:473–83. [PubMed: 16829979]
12. Smith E, Shilatifard A. The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. Mol Cell. 2010; 40:689–701. [PubMed: 21145479]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

13. Strahl BD, Allis CD. The language of covalent histone modifications. Nature. 2000; 403:41–5. [PubMed: 10638745]
14. Vermeulen M, et al. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell. 2007; 131:58–69. [PubMed: 17884155]
15. Herz HM, et al. Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes Dev. 2012; 26:2604–20. [PubMed: 23166019]
16. Hu D, et al. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Mol Cell Biol. 2013; 33:4745–54. [PubMed: 24081332]
17. Lee JE, et al. H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. Elife. 2013; 2:e01503. [PubMed: 24368734]
18. Outchkourov NS, et al. Balancing of histone H3K4 methylation states by the Kdm5c/SMCX histone demethylase modulates promoter and enhancer function. Cell Rep. 2013; 3:1071–9. [PubMed: 23545502]
19. Cheng J, et al. A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. Mol Cell. 2014; 53:979–92. [PubMed: 24656132]
20. Bartke T, et al. Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell. 2010; 143:470–84. [PubMed: 21029866]
21. Vermeulen M, et al. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell. 2010; 142:967–80. [PubMed: 20850016]
22. Carruthers LM, Tse C, Walker KP 3rd, Hansen JC. Assembly of defined nucleosomal and chromatin arrays from pure components. Methods Enzymol. 1999; 304:19–35. [PubMed: 10372353]
23. Luger K, Rechsteiner TJ, Richmond TJ. Expression and purification of recombinant histones and nucleosome reconstitution. Methods Mol Biol. 1999; 119:1–16. [PubMed: 10804500]
24. Luger K, Rechsteiner TJ, Richmond TJ. Preparation of nucleosome core particle from recombinant histones. Methods Enzymol. 1999; 304:3–19. [PubMed: 10372352]
25. Simon MD, et al. The site-specific installation of methyl-lysine analogs into recombinant histones. Cell. 2007; 128:1003–12. [PubMed: 17350582]
26. Engelen E, et al. Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. Nat Commun. 2015; 6:7155. [PubMed: 25990348]
27. Kagey MH, et al. Mediator and cohesin connect gene expression and chromatin architecture. Nature. 2010; 467:430–5. [PubMed: 20720539]
28. Dreveny I, et al. The double PHD finger domain of MOZ/MYST3 induces alpha-helical structure of the histone H3 tail to facilitate acetylation and methylation sampling and modification. Nucleic Acids Res. 2014; 42:822–35. [PubMed: 24150941]
29. Yue F, et al. A comparative encyclopedia of DNA elements in the mouse genome. Nature. 2014; 515:355–64. [PubMed: 25409824]
30. Li Y, et al. CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells. PLoS One. 2014; 9:e114485. [PubMed: 25486255]
31. Wang C, et al. Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. Proc Natl Acad Sci U S A. 2016; 113:11871–11876. [PubMed: 27698142]
32. Dorighi KM, et al. Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. Mol Cell. 2017; 66:568–576 e4. [PubMed: 28483418]
33. Phelan ML, Sif S, Narlikar GJ, Kingston RE. Reconstitution of a core chromatin remodeling complex from SWI/SNF subunits. Mol Cell. 1999; 3:247–53. [PubMed: 10078207]
34. Zeng L, et al. Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. Nature. 2010; 466:258–62. [PubMed: 20613843]
35. Li W, Zhao A, Tempel W, Loppnau P, Liu Y. Crystal structure of DPF3b in complex with an acetylated histone peptide. J Struct Biol. 2016; 195:365–72. [PubMed: 27402533]
36. Xiong X, et al. Selective recognition of histone crotonylation by double PHD fingers of MOZ and DPF2. Nat Chem Biol. 2016; 12:1111–1118. [PubMed: 27775714]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

37. Ho L, Crabtree GR. Chromatin remodelling during development. Nature. 2010; 463:474–84. [PubMed: 20110991]

38. Liu X, et al. In Situ Capture of Chromatin Interactions by Biotinylated dCas9. Cell. 2017; 170:1028–1043 e19. [PubMed: 28841410]

39. Rickels R, et al. Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. Nat Genet. 2017; 49:1647–1653. [PubMed: 28967912]

1. Luger K, Rechsteiner TJ, Richmond TJ. Expression and purification of recombinant histones and nucleosome reconstitution. Methods Mol Biol. 1999; 119:1–16. [PubMed: 10804500]

2. Luger K, Rechsteiner TJ, Richmond TJ. Preparation of nucleosome core particle from recombinant histones. Methods Enzymol. 1999; 304:3–19. [PubMed: 10372352]

3. Simon MD, et al. The site-specific installation of methyl-lysine analogs into recombinant histones. Cell. 2007; 128:1003–12. [PubMed: 17350582]

4. Dyer PN, et al. Reconstitution of nucleosome core particles from recombinant histones and DNA. Methods Enzymol. 2004; 375:23–44. [PubMed: 14870657]

5. Carruthers LM, Tse C, Walker KP 3rd, Hansen JC. Assembly of defined nucleosomal and chromatin arrays from pure components. Methods Enzymol. 1999; 304:19–35. [PubMed: 10372353]

6. Carey MF, Peterson CL, Smale ST. Dignam and Roeder nuclear extract preparation. Cold Spring Harb Protoc. 2009; 2009 pdb prot5330.

7. Dignam JD, Martin PL, Shastry BS, Roeder RG. Eukaryotic gene transcription with purified components. Methods Enzymol. 1983; 101:582–98. [PubMed: 6888276]

8. Kapust RB, Waugh DS. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci. 1999; 8:1668–74. [PubMed: 10452611]

9. Carey MF, Peterson CL, Smale ST. In vivo DNase I, MNase, and restriction enzyme footprinting via ligation-mediated polymerase chain reaction (LM-PCR). Cold Spring Harb Protoc. 2009; 2009 pdb prot5277.

10. Albuquerque CP, et al. A multidimensional chromatography technology for in-depth phosphoproteome analysis. Mol Cell Proteomics. 2008; 7:1389–96. [PubMed: 18407956]

11. Chen SH, Albuquerque CP, Liang J, Suhandynata RT, Zhou H. A proteome-wide analysis of kinase-substrate network in the DNA damage response. J Biol Chem. 2010; 285:12803–12. [PubMed: 20190278]

12. Lowary PT, Widom J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol. 1998; 276:19–42. [PubMed: 9514715]

13. Cao XJ, Arnaudo AM, Garcia BA. Large-scale global identification of protein lysine methylation in vivo. Epigenetics. 2013; 8:477–85. [PubMed: 23644510]

14. Yun M, Ruan C, Huh JW, Li B. Reconstitution of modified chromatin templates for in vitro functional assays. Methods Mol Biol. 2012; 833:237–53. [PubMed: 22183598]

15. Wang C, et al. Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. Proc Natl Acad Sci U S A. 2016; 113:11871–11876. [PubMed: 27698142]

16. Kim TH, et al. Direct isolation and identification of promoters in the human genome. Genome Res. 2005; 15:830–9. [PubMed: 15899964]

17. Li Z, et al. A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. Proc Natl Acad Sci U S A. 2003; 100:8164–9. [PubMed: 12808131]

18. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. [PubMed: 19261174]

19. Dorighi KM, et al. Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. Mol Cell. 2017; 66:568–576 e4. [PubMed: 28483418]

20. Kabsch W. Xds. Acta Crystallogr D Biol Crystallogr. 2010; 66:125–32. [PubMed: 20124692]

21. Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr. 2013; 69:1204–14. [PubMed: 23793146]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

22. Winn MD, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr. 2011; 67:235–42. [PubMed: 21460441]
23. Terwilliger TC, et al. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. Acta Crystallogr D Biol Crystallogr. 2009; 65:582–601. [PubMed: 19465773]
24. McCoy AJ, Storoni LC, Read RJ. Simple algorithm for a maximum-likelihood SAD function. Acta Crystallogr D Biol Crystallogr. 2004; 60:1220–8. [PubMed: 15213383]
25. McCoy AJ, et al. Phaser crystallographic software. J Appl Crystallogr. 2007; 40:658–674. [PubMed: 19461840]
26. Terwilliger TC. Maximum-likelihood density modification. Acta Crystallogr D Biol Crystallogr. 2000; 56:965–72. [PubMed: 10944333]
27. Terwilliger TC. Automated main-chain model building by template matching and iterative fragment extension. Acta Crystallogr D Biol Crystallogr. 2003; 59:38–44. [PubMed: 12499537]
28. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010; 66:486–501. [PubMed: 20383002]
29. Afonine PV, et al. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr. 2012; 68:352–67. [PubMed: 22505256]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

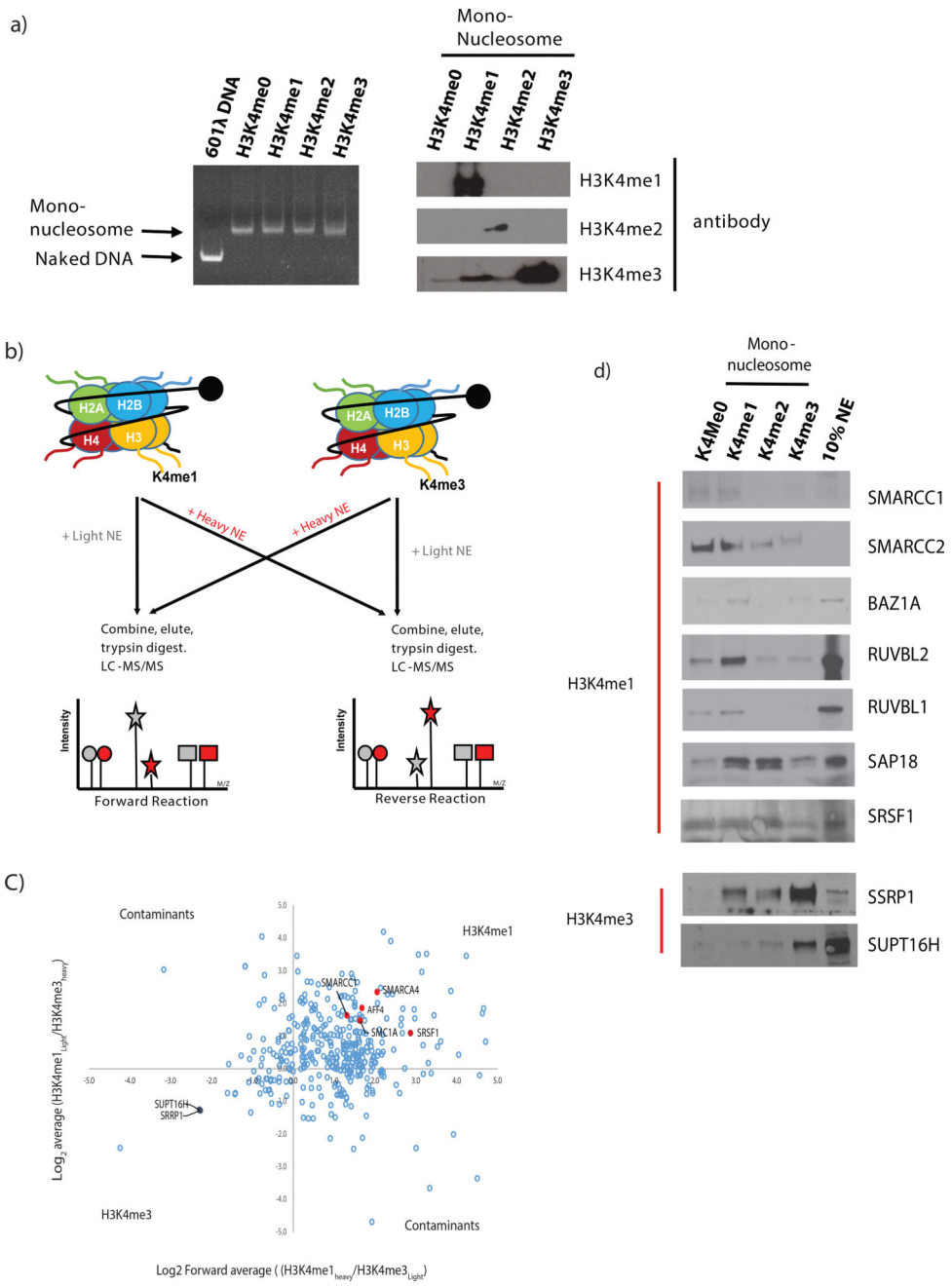

Figure 1. Identification of H3K4me1 binding proteins using SILAC and Mass-spec analysis

A) Left – Mononucleosomes assembled from biotin tagged 601λ positioning sequence and methylated octamers. Right – Chemically modified nucleosomes are recognized by specific antibodies against various H3K4 methylation. 3 independent chemical modifications were tested yielding similar results. B) Schematic of SILAC mass spec screen. C) Average Log2 L/H of forward reactions on X-axis and log2 H/L of reverse reactions on y-axis (from 4 independent biological replicates). Top right quadrant is H3K4me1 associated factors and bottom left quadrant contains H3K4me3 associated factors. D) Biotin-tagged methylated nucleosomes used as bait for pulldowns from HeLa NE. The bound proteins detected by

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

western blotting with specific antibodies are listed, experiments were repeated at least twice with similar results.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

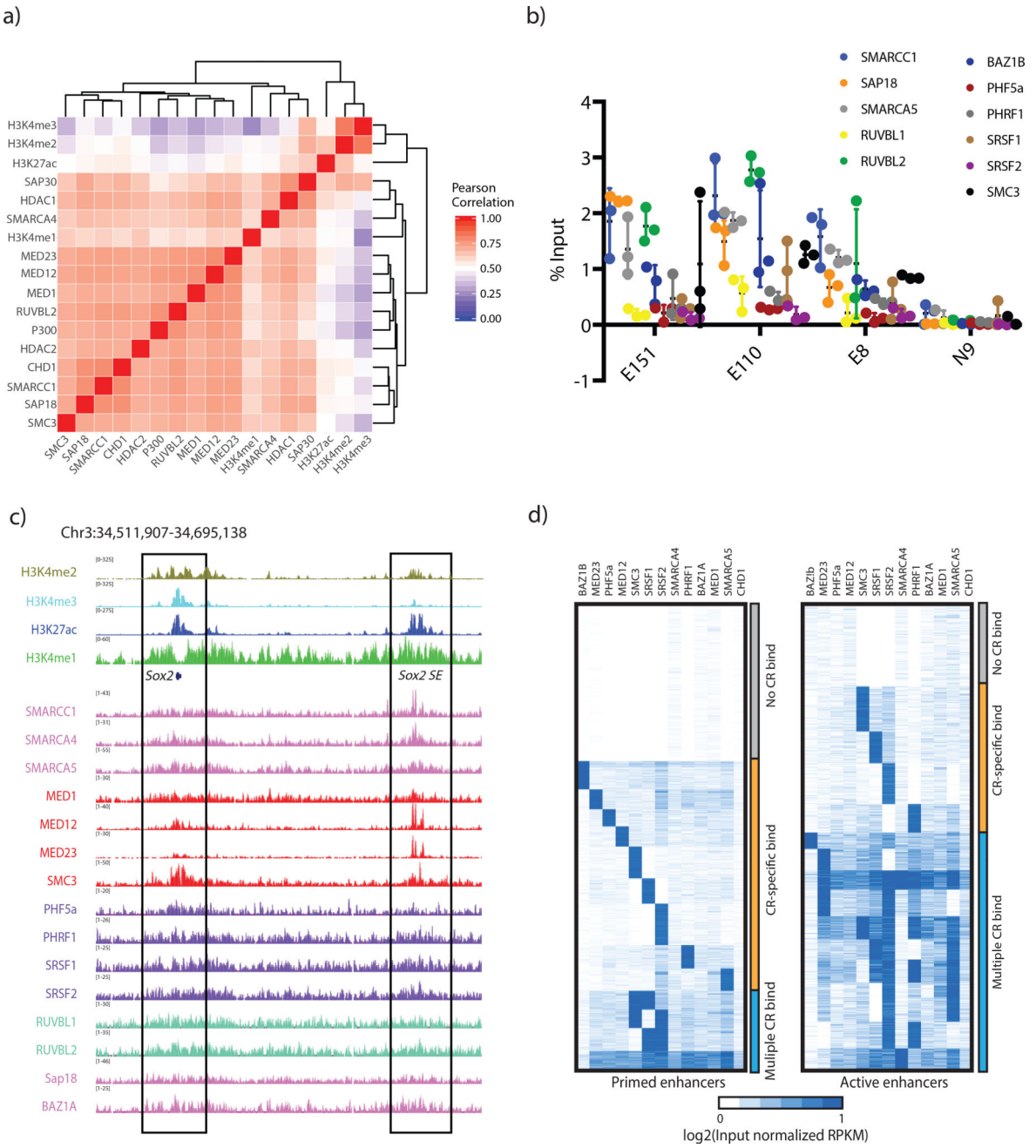

Figure 2. Binding of CRs at H3K4me1 regions and enhancers

A) Hierarchical clustering of genome-wide ChIP-seq signals (RPKM) for H3K4me1, H3K4me2, H3K4me3 and chromatin binding proteins with 1kb-binning, n=2,435,743. The heatmap shows pair-wise Pearson correlation coefficient between different ChIP-seq datasets. B) ChIP-qPCR in mESC with antibodies listed, primers designed for validated enhancers E110, E151, E8 and negative control region N9. Error bars, mean ±SD for n=3 biological replicates. C) Browser shot of candidate H3K4me1 readers at the Sox2 enhancer. Active enhancer with high H3K27ac boxed left, poised enhancer with low H3K27ac boxed right. D) Heat maps for K-means clustering results of input normalized CR signals

according to poised enhancers versus active enhancers. Each cluster was manually classified as ‘Multiple CR bind’, ‘CR-specific bind’, and ‘No CR bind’ according to CR binding patterns. Experiments were repeated at least twice with each antibody.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

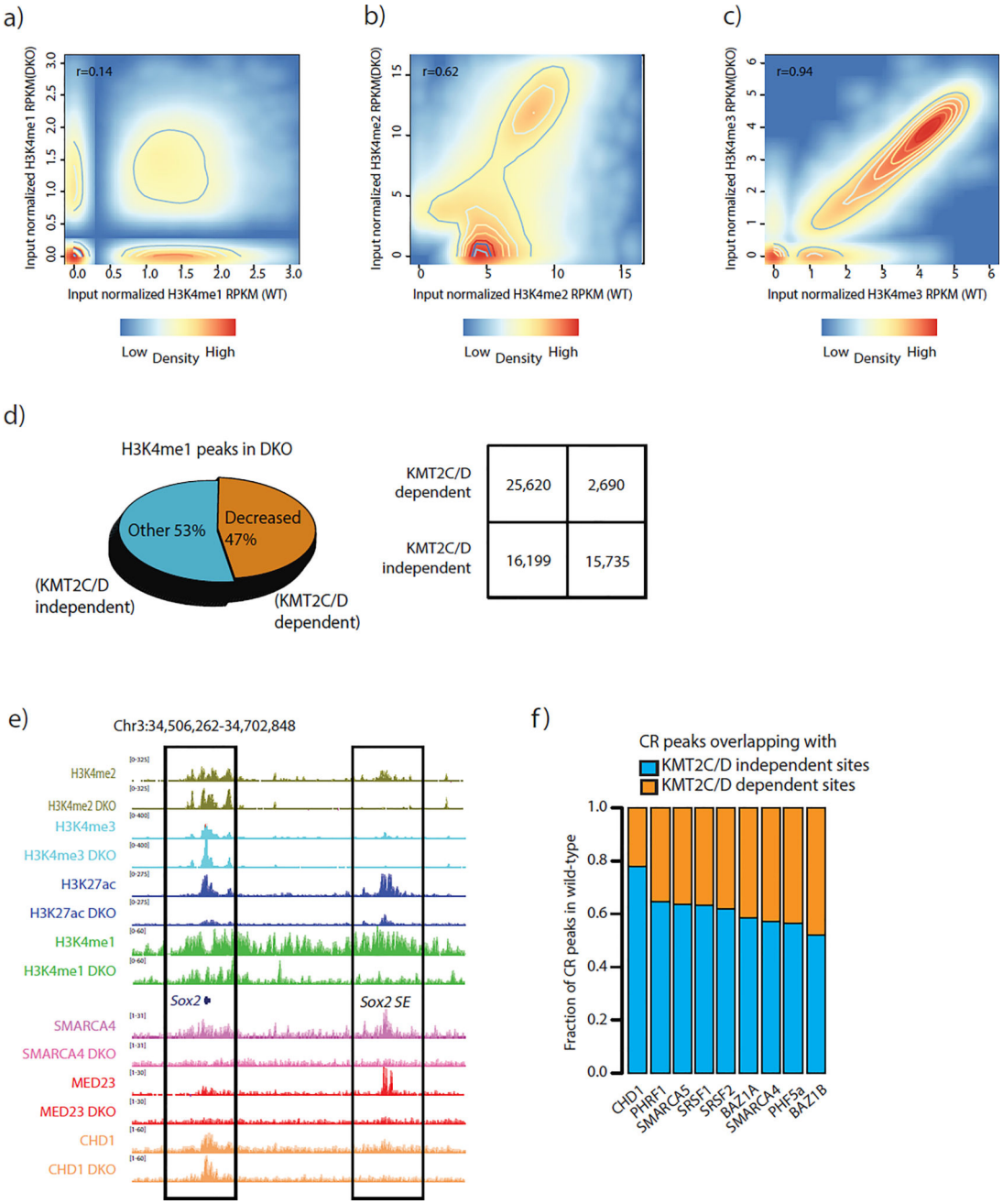

Figure 3. Concomitant loss of H3K4me1 and CR binding at enhancers in KMT2C/D DKO mouse ES cells

A) A scatter density plot of input normalized H3K4me1 RPKMs between wild-type and KMT2C/D DKO cell lines at H3K4me1 peaked regions, n=43,918. B) A scatter density plot of input normalized H3K4me2 RPKMs between WT and KMT2C/D DKO cell lines at H3K4me2 peaked regions, n=33,197. C) A scatter density plot of input normalized H3K4me3 RPKMs between wild-type and KMT2C/D DKO cell lines at H3K4me3 peaked regions, n=22,157. D) Upper panel - A pie chart for the fraction of H3K4me1 peaks in DKO KMT2C/D mESCs according to KMT2C/D dependent or KMT2C/D independent patterns. Lower panel - 2 by 2 table of the relationship with enhancer regions according to KMT2C/D

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

dependent and independent H3K4me1 peaked regions. E) Browser shot of H3K4me1, H3K4me2, H3K4me3, H3K27ac, and CR levels in WT vs DKO KMT2C/D mESCs at the Sox2 enhancer. For each factor top track in form WT and bottom track is DKO. F) Bar plots are shown for the fraction of CR peaks in wild-type (y-axis) according to overlap with KMT2C/D independent (blue) and dependent sites (orange). Total number of CR peaks identified are : CHD1 (n=14,846), PHRF1 (n=21,924), SMARCA5 (n=13,891), SRSF1 (n=23,221), SRSF2 (n=31,200), BAZ1A (n=13,806), SMARCA4 n=10,897), PHRF5a (n=11,926), BAZ1B (n=5,405). Experiments were repeated at least twice in each cell type.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

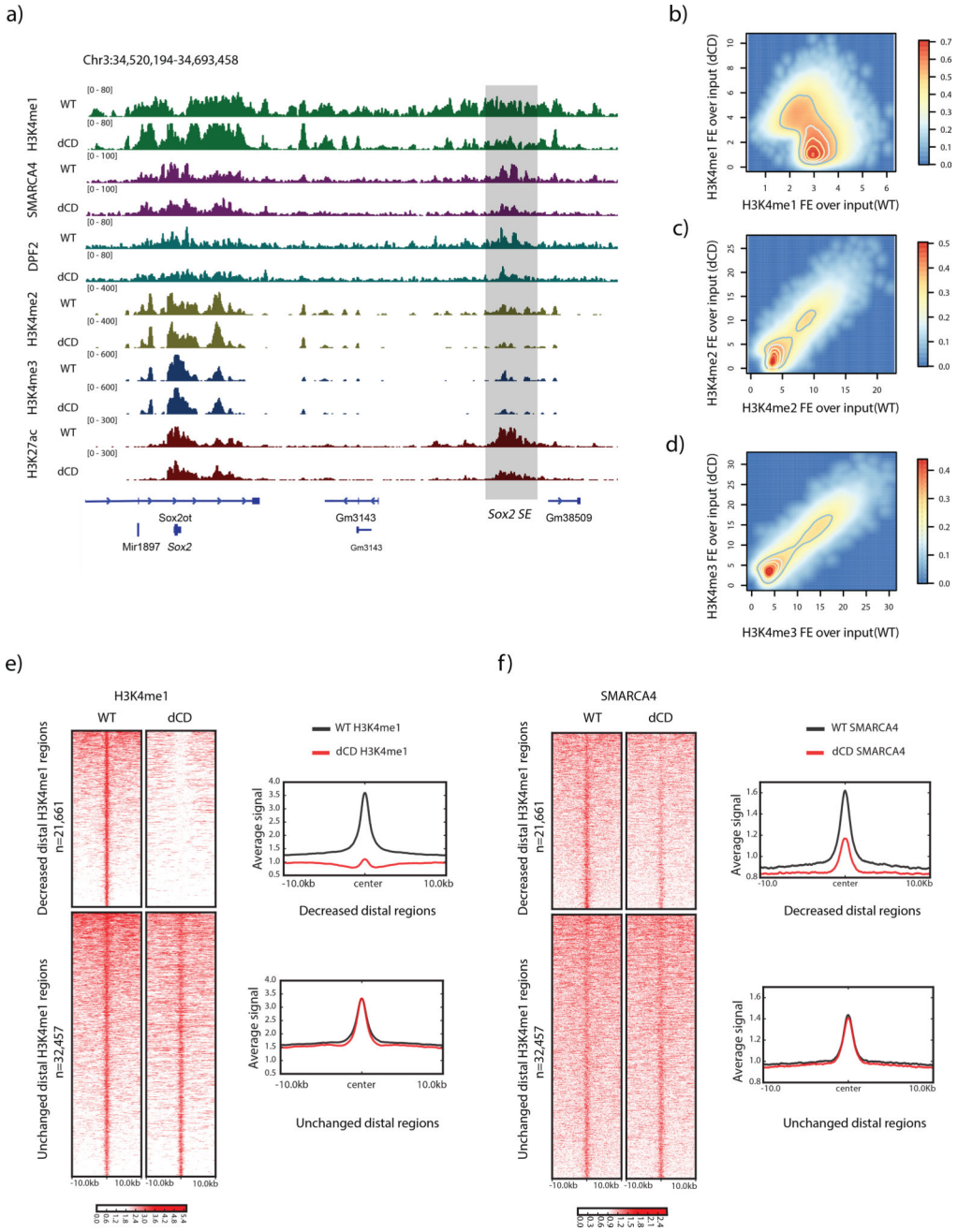

Figure 4. Reduced BAF complex binding is associated with depletion of H3K4me1 in KMT2C/D catalytically null (dCD) cells

A) Browser shot of ChIP-seq signal (RPKM) for SMARCA4 and DPF2 at the Sox2 locus. The Sox2 super-enhancer is shaded on right. Experiments were repeated independently twice with similar results. B,C,D) Scatter density plots of input normalized fold enrichment between WT and dCD at H3K4me1(n=82,053), H3K4me2(n=53,501) and H3K4me3(n=34,553) peaked regions. E) Left - Heatmap of input normalized H3K4me1 ChIP signal in WT and dCD over 21,661 distal H3K4me1 regions with decreased signals in dCD and 32,475 distal H3K4me1 regions with invariable signals, with regions sorted by strength of H3K4me1 signal. Right - aggregate plot showing the average signal in WT and

dCD. F) Left - Heatmap of input normalized SMARCA4 ChIP-seq signal in WT and DCD over the same regions in E. Right - aggregate plot showing the average signal in WT and dCD.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

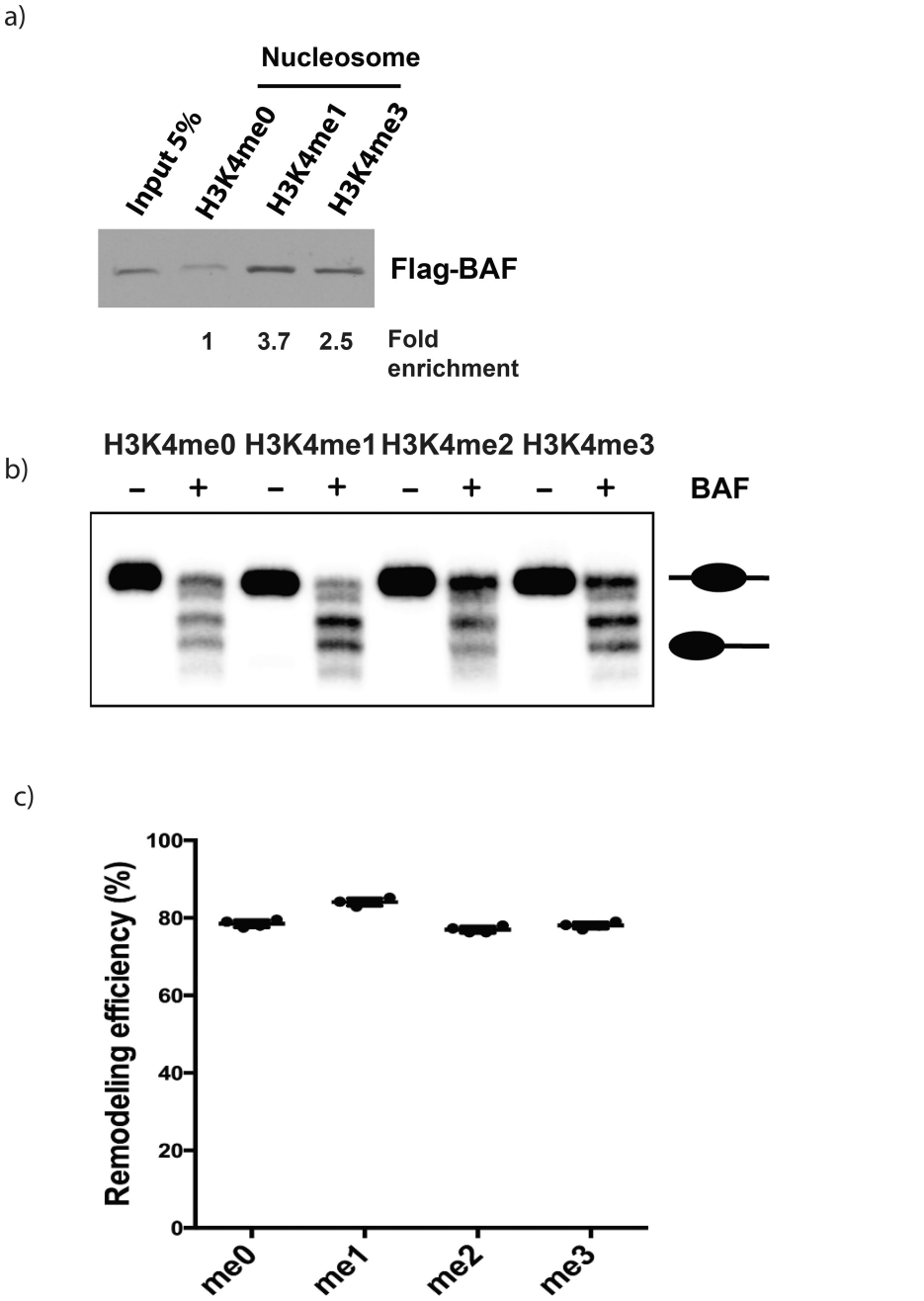

Figure 5. BAF complex preferentially binds and remodels H3K4me1 modified nucleosomes

A) Purified Flag-BAF complex binding to H3K4 methylated-nucleosomes, western blotted with anti-FLAG antibody (M2). Pulldown repeated 3 times yielding the same result. B) Polyacrylamide gel showing representative (n=4) *in vitro* remodeling assay. After incubation with BAF complex, nucleosomes are slid to the end of the 216-bp DNA fragment resulting in a change in mobility in the gel. Top band is un-remodeled nucleosome, and lower four bands are slid nucleosomes with different positions away from 146-bp Widom601 binding sites in the middle. C) Quantification of nucleosome remodeling assays. Error bars, mean

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

±SD n=4 biological replicates, see Figure S4C. The reduced percentage of the top band is defined as remodeling efficiency.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

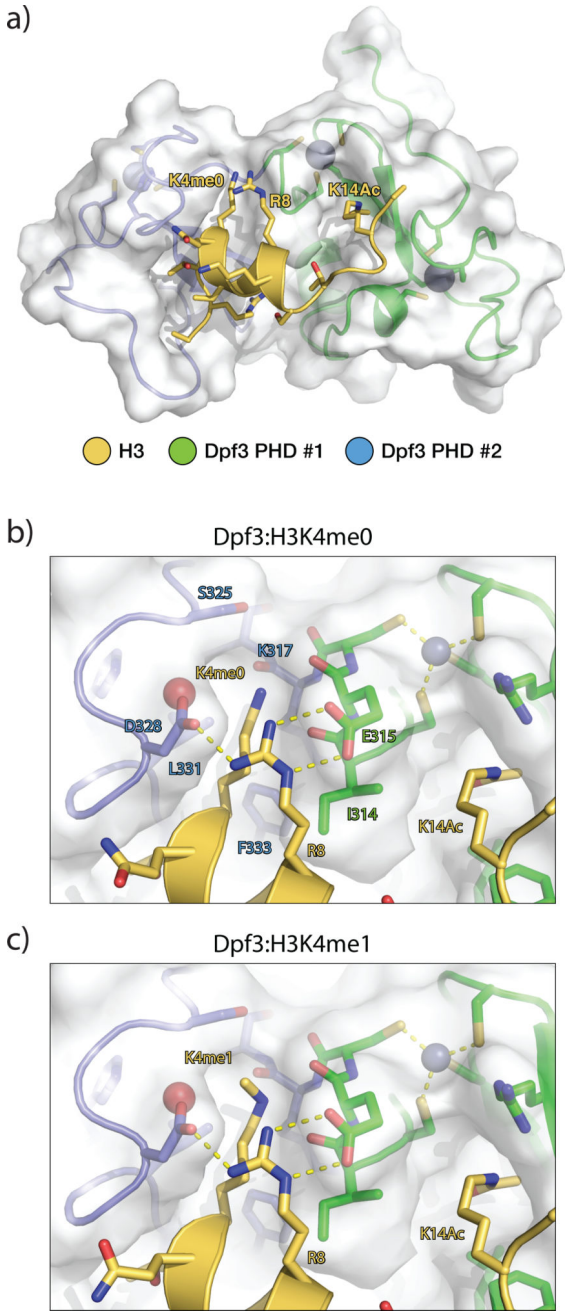

Figure 6. Structural basis for H3K4 recognition by DPF3

A) Overall structure of DPF3:H3K4me0 complex. DPF3 PHD1 domain is shown in green, PHD2 in blue, and histone H3 tail peptide shown in yellow. B) Close-up view of the DPF3 PHD1–2 region (light blue, white surface) with H3 residues 1–18 with H3K4me0 and H3K14ac (yellow). PHD1 binds H3K14ac as previously observed, while PHD2 binds H3K4 and H3R8. C) Close-up view of DPF3 binding H3 1–18 with H3K4me1 and H4K14ac. The mono-methyl group is accommodated in a pre-formed surface pocket on DPF3. For views of the overall structure and electron density maps, see Figure S5.